GTSPP Meeting (September 27, 2004)

2.Data Processing

We, at AOML encountered the following problems when we processed the 1999 Delayed mode data sets:

a) Duplicate Profiles:

Since taking over qc for XBT's, 10 years ago, duplicate profiles have been a major problem. The 1999 Delayed mode data sets showed different kinds of dups as compared to the previous years. For example, between 1990 through 1998, the majority of dups were close dups. In other words, they were very close in time, location, and/or depth/temp values. In 1999, dups were a mixture of close dups and wrong sign dups. I observed that the 1999 data set has dups with the same date and only a different longitude or latitude sign. For example, one profile has a latitude of 0.667 and a longitude of 8.367 and the other profile has a latitude of -0.670 and a longitude of 8.370. Also, the number of lists under "SRFC_CODES" and the history records were different but amazingly, their depths and temperatures were identical.   It was difficult to decide which profile was the duplicated one for a wrong sign dup case. Obviously, one has the wrong sign location. I forwarded these cases back to Melanie and let her identify the dups from her original profiles.

When we get a delayed mode dataset, the first thing we do is run our dup test. After the possible dup lists are created we then run a visual dup test to verify and mark them as dup profiles. This dup information is stored in the "DUP_FLAG" field in MEDS_ASCII format, but there is no field in the MEDS_ASCII format that holds the primary (i.e. dup of which profile) profile information. The majority of dups were close or near dups. We have saved the dup files that show a list of primary profile id with the duplicate profile id. I do not know if NODC is using this "DUP_FLAG" to mark the profiles in their databases as dup. I would say between 6-10% of the profiles were identified as dups. I believe that all these dups were a result of human or computer errors along the way. One possibility might be that the same profile was sent more than once to GTS. It could have been a transmission error while transmitting the data to GTS. It could have been programming mistakes loading the same profile more than once to the database. It could have been rounding errors from one machine to another that treat the same profiles as two different profiles.

I have downloaded the best copy of GTSPP data from the NODC web page and found that the duplicates and flags found by AOML were not included in the best copy of the dataset.
The best copy did not have the dup flags set for those profiles that Melanie had identified as dups for the wrong sign location cases. One example is DBID=1315288 from the best copy of "atq11999". It should be set as dup but it is not.
I have checked the entire best dataset from 1990 to 1999 and none of AOML's duplicate identifiers were included in the best datasets.

b) GTSPP Operation vs. ARGO Operation

Unlike the ARGO operation, GTSPP MEDS_ASCII format files are not easily replaceable by the data originator. The MEDS_ASCII format is not user friendly. NetCDF files in US GODAE and IFREMER are much more user friendly and self-explainitory.

The ARGO community is receiving a dup list report once a month so the originator can fix the problem immediately and send a corrected NetCDF file to US GODAE & IFREMER. I think that the three science centers (AOML, CSIRO and SIO) and two data ceneters (NODC & MEDS) need to agree on the criteria of the duplicate definitions and how to update the best copy datasets.

c) Data type ="DT" (Digital BT) problem

I had not received any 'DATA_TYPE'='DT' prior to the 1999 delayed mode dataset. I also learned that the 'DT' comes from the XBT's that were recorded on a digital logger. I changed my program to accept the 'DT' as one of our acceptable data types starting with the 1999 delayed mode dataset. When did this 'DT' become introduced in to the GTSPP? AOML lost all profiles with 'DATA_TYPE'='DT' prior to 1999 dataset.

d) History record "PREV_VAL" is now character string

I noticed that when 'prev_act'='PLAT', 'Prev_val' is no longer numeric. The 'Prev_val' can now be a character string as well.

5.Status report on the use of the CRC tag

According to Steve Cook, by the end of this year all our XBT equipped VOS (about 40 ships) will be using SEAS2000 and the CRC tag. The following message is prepared by Paul Chinn who works in Silver Spring, MD on the SEAS2000 program:

There are actually 2 CRC's.   At present, for each SEAS 2000 bathy message received at the AMVER/SEAS (A/S) server in Silver Spring, a WMO coded bathy message is created for distribution on the GTS. Before the JJVV message is placed on the GTS, a real-time archive file is created for delivery to NODC. Included in the real-time archive is a 32 bit GTS CRC. This GTS CRC is computed on everything after the 888 group up to the and including the call sign. Theoretically, this GTS CRC should uniquely identify each JJVV message placed on the GTS from SEAS 2000. When MEDS receives the JJVV message on the GTS, they compute the CRC and place in the the GTSPP archive records they send to NODC. Now NODC has a record from MEDS and a record from us where the GTS CRCs should be equal. This helps identify duplicate messages.

The second CRC is the SEAS ID which is the 32 bit CRC of the raw SEAS binary message stored on the ship. This 32 bit CRC is transmitted to the AMVER/SEAS server in the binary message regardless of the type of bathy data in the message (flexure points, 2 meter, or full resolution). Once the message is received at the A/S a GTS bulletin and the real-time archive file are produced. The real-time archive file contains both the GTS CRC and the SEAS ID. The SEAS ID allows detection of different products generated from the same message. For example, if the ship rider opened a full resolution message and transmitted a binary message containing flexure points and later retransmitted the same message containing 2 meter data, two real-time archive files would be sent to NODC with different GTS CRCs but the same SEAS ID indicating that the flexure point bulletin and the 2 meter originated from the same delayed mode message. (The original full res file/message stays on the ship until archived and collected and therefore is delayed mode data). When we get the delayed mode data and create delayed mode archive files for submission to NODC, the SEAS ID identifies what records require replacing.

I'm sure this is clear as mud but don't hesitate to call. But in answer to your question, MEDS, NODC and us are still kind of figuring out how its working. I need to draw a data flow diagram.

I've attached one of the real-time archive files.

Paul

6.Links to other programs

I would like to report one case that I encountered . I let Melanie Hamiton at NODC know about one ARGO float that had been gone in to the GTS, about 10 times, with a wrong salinity. This problem float's wmo id happened to be 3900118. She told me that she will notify the MEDS. My concern is this. We already performed the real-time QC on this profile from ARGO process and this float will be feed into our GTSPP real-time data stream and later it will get QC'ed by the Science center again. Why does GTSPP quality control ARGO data again and how can we prevent it?